3 Chi-Squarred Method for Recommendation

3.1 χ^2 Applied to Common Reviewers

This recommendation method is based on the χ^2 statistic for independence of rows and columns in a contingency table. The main idea is to generate a contingency table based on common reviewers of two particular movies. This is constructed based on a binary variable **liked**, which can take the value L (liked movie) and N (not liked movie). An L is assigned for review scores equal to 5, and N is assigned otherwise. The contingency table is created by crossing this variable for movie i and movie j.

```
contingency.table <- table(liked.j,liked.i)
contingency.table
```

```
##         liked.i
## liked.j  L  N
##       L 13  5
##       N  4 11
```

```
chisq.test(contingency.table)$statistic
```

```
## X-squared
##  5.096599
```

The purpose of the χ^2 is to test the null hypothesis that rows are independent from columns. The statistic is computed by summing the squared differences between the observed and the expected count assuming independence. Therefore, under the assumption of independence the observed should be close to the expected and a small χ^2 should be observed. The following is an example of this:

```
contingency.table <- table(liked.j,liked.i)
contingency.table
```

```
##         liked.i
## liked.j  L  N
##       L  3  7
##       N  4 10
```

```
chisq.test(contingency.table)$statistic
```

```
##    X-squared
## 1.471656e-31
```

For cases with a strong row and column dependence, a high value of the χ^2 is expected. It is important to

note that strong dependence could have two meanings. The first one, that users who reviewed both movies agree most of the time, i.e., the majority either liked both or disliked both. The first example that was shown is a case of this. There is a second case of strong dependence where the users most of the time disagree, i.e., they liked one of the movies but disliked the other. The following is an example of this case:

```
contingency.table <- table(liked.j,liked.i)
contingency.table
```

```
##         liked.i
## liked.j  L  N
##       L  4  9
##       N 12  2
```

```
chisq.test(contingency.table)$statistic
```

```
## X-squared
##  6.306842
```

Note that both examples show a large χ^2 , however they have a different interpretation. For the purpose of this recommendation tool, only case one was considered. Even though case two is also meaningful, it was not considered in the sake of simplicity. The χ^2 was computed only for cases where the following is true:

$$LL + NN > NL + LN$$

3.2 Recommendation Tool

The recommendation tool is currently implemented as an R function, however it could easily be implemented in a shiny app. χ^2 values were computed as described before for each pair of movies in the training sample and saved in a data frame called PairChi. The user should give as a first input the movie id of one of their favorite movies. The second input is n, the number of recommnedations that want to be retreived. The function will extract from PairChi the χ^2 values corresponding to the input movie and the rest of the movies, they will be ranked and then a function will eliminate repetions (for example if there's a Titanic VHS and Titanic DVD only the first one will be considered). The first n movies will be recommended to the user. The following is an example of this:

```
return.name("B00004TX12")
chi.sqr.recommendation("B00004TX12",5)
```

```
## [1] "Watchmen Collector's Edition: Ultimate Cut + Graphic Novel [Blu-ray]"
```

```
##                                                 name2      chi
## 88                          Superman Returns [Blu-ray] 3.932664
## 107                            V for Vendetta [HD DVD] 2.387153
## 73  Sin City - Unrated (Two-Disc Collector's Edition) 2.343750
```